International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:4, No:4, 2010

Application of Adaptive Neuro-Fuzzy Inference System in Smoothing Transition Autoregressive Models

E. Giovanis

Abstract**—** In this paper we propose and examine an Adaptive Neuro-Fuzzy Inference System (ANFIS) in Smoothing Transition Autoregressive (STAR) modeling. Because STAR models follow fuzzy logic approach, in the non-linear part fuzzy rules can be incorporated or other training or computational methods can be applied as the error backpropagation algorithm instead to nonlinear squares. Furthermore, additional fuzzy membership functions can be examined, beside the logistic and exponential, like the triangle, Gaussian and Generalized Bell functions among others. We examine two macroeconomic variables of US economy, the inflation rate and the 6-monthly treasury bills interest rates.

Keywords**—** Forecasting, Neuro-Fuzzy, Smoothing transition, Time-series

I. INTRODUCTION

EMPIRICAL analysis in macroeconomics as well as in financial economics is largely based on times series. This approach allows the model builder to use statistical inference in constructing and testing equations that characterize relationships between economic variables. There are two kinds of econometric modelling in time-series analysis. The first one contains the linear models like Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models among other. The second is consisted b non-linear models, as the Threshold Autoregressive (TAR) models, Smoothing Transition Autoregressive (STAR) Models and Markov Switching Regime Autoregressive (MS-AR) model.

One criticism in STAR modeling is that the estimation procedure can be incomplete. To be specific the linear part is exactly like a linear Autoregressive (AR) process. But the non-linear part is actually a fuzzy database, where the values are the fuzzy membership grades of the inputs. So a first notice is that no rules and no linguistic terms are introduced and for inputs more than one the AND-OR operators are not considered. The second criticism is that the nonlinear squares with Levenberg-Marquardt algorithm might not be appropriate. Therefore taking the fuzzy rules and other optimization procedures like linear programming or neuro-fuzzy approach with error backpropagation algorithm can be more efficient optimization techniques. One of the very few studies which approximated that is the study of Aznarte *et al.* [1], who have proved that the smooth transition autoregressive (STAR) model is functionally equivalent to a restricted fuzzy rule-based system, implying that the tools and theoretical results developed in each one of the two areas can be applied to the other area as well.

The purpose of this paper is to propose a neuro-fuzzy approach in STAR modeling. Fuzzy logic is an effective rule-based modelling in soft computing, that not only tolerates imprecise information, but also makes a framework of approximate reasoning. The disadvantage of fuzzy logic is the lack of self learning capability. On the other hand neural networks are capable to describe non-linearities, but are considered as black-boxes. The combination of fuzzy logic and neural network can overcome the disadvantages of the above approaches. In ANFIS, is combined both the learning capabilities of a neural network and reasoning capabilities of fuzzy logic in order to give enhanced prediction capabilities.

ANFIS has been used by many researchers to forecast various time Series comparing with Autoregressive (AR) and Autoregressive Moving Average (ARMA) models finding superior results in favour of ANFIS [2]-[4].

In section II we present the methodology of STAR models and neuro-fuzzy approach used in this study, while in section III and IV we present the data and the empirical results.

II. METHODOLOGY

A. Smoothing Transition Autoregressive Models

The smoothing transition auto-regressive (STAR) model was introduced and developed by Chan and Tong [5] and is defined as:

$$y_t = \pi_{10} + \pi_1{}' w_t + (\pi_{20} + \pi_2{}' w_t) F(y_{t-d}; \gamma, \mathrm{c}) + u_t \quad (1)$$

,where $u_t \sim (0,\sigma^2)$, π_{10} and π_{20} are the intercepts in the middle (linear) and outer (nonlinear) regime respectively, $w_t = (y_{t-1} \ldots y_{t-j})$ *is* the vector of the explanatory variables consisting of the

Eleftherios Giovanis is with the Royal Holloway University of London, department of Economics, Egham, Surrey TW20 0EX, UK, e-mail: giovanis@freemail.gr, Eleftherios.Giovanis.2010@live.rhul.ac.uk dependent variable with $j=1...p$ lags, y_{t-d} is the transition variable, parameter c is the threshold giving the location of the transition function and parameter γ is the slope of the transition function. The STAR model estimation is consisted by three steps according to Teräsvirta [6].

a) The specification of the autoregressive *(AR)* process of $j=1,\dots p$. One approach is to estimate *AR* models of different order and the maximum value of j can be chosen based on the Akaike (*AIC*) information criterion Besides this approach, j value can be selected by estimating the auxiliary regression (2) for various values of $j=1,\dots p$, and choose that value for which the *P-value* is the minimum, which is the process we follow.

b) The second step is testing linearity for different values of delay parameter d. We estimate the following auxiliary regression:

$$y_t = \beta_o + \beta_1 w_t + ...+ \sum_{j=1}^{p} \beta_{2j} y_{t-j} y_{t-d} + \sum_{j=1}^{p} \beta_{3j} y_{t-j} y_{t-d}^{2} + \sum_{j=1}^{p} \beta_{4j} y_{t-j} y_{t-d}^{3} + \varepsilon_t \quad (2)$$

The null hypothesis of linearity is H_0: $\beta_{2j} = \beta_{3j} = \beta_{4j} = 0$. In order to specify the parameter d the estimation of (2) is carried out for a wide range of values $1 \le d \le D$ and we choose $d=1,\dots,5$ In the cases where linearity is rejected for more than one values of d, then d is chosen by the minimum value of $p(d)$, where $p(d)$ is the *P-value* of the linearity test. We examine for $j=1,\dots, 5$ and we choose those values of j and d , where the *P-value* is minimized.

c) The third and last step is the specification of STAR model. We test the following hypotheses by [6]-[7]

$$H_{04} : \beta_{4j} = 0, j = 1,..., p \quad (3)$$

$$H_{03} : \beta_{3j} = 0 \mid \beta_{4j} = 0, j = 1,..., p \quad (4)$$

$$H_{02} : \beta_{2j} = 0 \mid \beta_{3j} = \beta_{4j} = 0, j = 1,..., p \quad (5)$$

If we reject the (3) hypothesis then we choose LSTAR model. If (3) is accepted and (4) is rejected then ESTAR model is selected. Finally accepting (3) and (4) and rejecting (5) we choose LSTAR model. We present the results of the STAR model chosen based on test hypotheses, but we examine both LSTAR and ESTAR models, to examine the forecasting performance and to show that the difference between their predicting performances can be small.

We shall consider two transition functions, the logistic and the exponential [7], which are defined by (6) and (7) respectively.

$$F(y_{t-d}) = (1+\exp[-\gamma(y_{t-d}-c)])^{-1}, \gamma > 0 \quad (6)$$

$$F(y_{t-d}) = 1 - \exp(-\gamma(y_{t-d}-c)^2), \gamma > 0 \quad (7)$$

We apply a grid search procedure for equation (1) with non linear squares and Levenberg-Marquardt algorithm.

B. Unit Root and Stationary Tests

It is possible that the variables are not stationary in the levels, but probably are in the first or second differences. To be specific we confirm this assumption by applying Augmented Dickey-Fuller-ADF [8] and KPSS stationary test [9]. The ADF test is defined from the following relation:

$$\Delta y_t = \mu + \gamma y_{t-1} + \phi_1 \Delta y_{t-1} + ... + \phi_p \Delta y_{t-p} + \beta t + \varepsilon_t \quad (8)$$

, where y_t is the variable we examine each time. In the right hand of (8) the lags of the dependent variable are added with order of lags equal with p. Additionally, regression (8) includes the constant or drift μ and trend parameter β. The disturbance term is defined as ε_t. In the next step we test the hypotheses:

H_0: $\varphi=1$, $\beta=0$ => $y_t \sim I(0)$ with drift

against the alternative

H_1: $|\varphi|<1$ => $y_t \sim I(1)$ with deterministic time trend

The KPSS statistic is then defined as:

$$KPSS = T^{-2} \sum_{t=1}^{T} s^2{}_t / \hat{\sigma}^2(p) \cdot \quad (9)$$

, where T is the number of sample and $\hat{\sigma}^2(p)$ is the long-run variance of ε_t and can be constructed from the residuals ε_t as:

$$\hat{\sigma}^2(p) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + \frac{2}{T}\sum_{t=1}^{p} w_j(p) + \sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}, \quad (10)$$

, where p is the truncation lag, $w_j(p)$ is an optional weighting function that corresponds to the choice of a special window [10]. Under the null hypothesis of level stationary,

$$KPSS \rightarrow \int_0^1 V_1(r)^2\,dx \quad (11)$$

, where $V_1(x)$ is a standard Brownian bridge: $V_1(r) = B(r) - rB(1)$ and $B(r)$ is a Brownian motion (Wiener process) on $r \in [0, 1]$. Because relation (11) is refereed in testing only on the intercept and not in the trend and as we are testing with both intercept and trend we have the second-level Brownian bridge $V_2(x)$ and it is:

$$KPSS \rightarrow \int_0^1 V_2(r)^2\,dx \quad (12)$$

, where $V_2(x)$ is given by:

$$V_2(r) = W(r) + (2r - 3r^2)W_1 + (-6r + 6r^2)\int_0^1 W_s(s)ds \quad (13)$$

C. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Jang [11] and Jang and Sun [12] introduced the adaptive neuro-fuzzy inference system (ANFIS). This system makes use of a hybrid learning rule to optimize the fuzzy system parameters of a first order Sugeno system. An example of a two input with two rules first order Sugeno system can be graphically represented by figure 1.

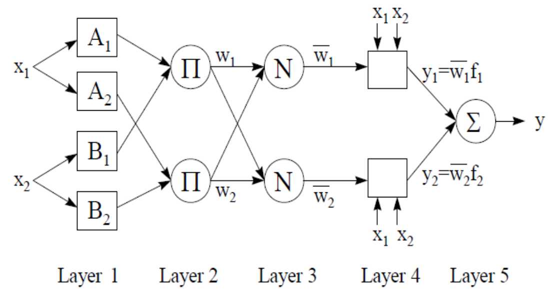

Fig. 1. Example of ANFIS architecture for a two-input, two-rule first-order Sugeno model

The ANFIS architecture is consisted of two trainable parameter sets, the antecedent membership function parameters and the polynomial parameters *p,q,r*, also called the consequent parameters. The ANFIS training paradigm uses a gradient descent algorithm to optimize the antecedent parameters and a least squares algorithm to solve for the consequent parameters. Because it uses two very different algorithms to reduce the error, the training rule is called a hybrid. The consequent parameters are updated first using a least squares algorithm and the antecedent parameters are then updated by backpropagating the errors that still exist. We define three linguistic terms {low, medium, high}. In the case where we accept that there is an AR(1) process in STAR models then we have one input; the dependent variable with one lag. In that case we do not take the *AND-OR* operators. Besides that we show also an example with two inputs.

The rules in the case of one input and with three linguistic terms are:

IF y_{t-1} is low THEN $f_1 = p_1x + r_1$

IF y_{t-1} is medium THEN $f_2 = p_2x + r_2$

IF y_{t-1} is high THEN $f_3 = p_3x + r_3$

, where y_{t-1} is the dependent variable with one lag. In that case the steps for ANFIS will be:

$$O_i^1 = \mu_{A_i}(x) \quad (14)$$

The adjustable parameters that determine the positions and shapes of these node functions are referred to as the premise parameters. In layer 2 we have:

$$O_i^2 = w_i = \prod_{j=1}^{m} \mu_{A_i}(x) \quad (15)$$

Each node output represents the firing strength of the reasoning rule. In layer 3, each of these firing strengths of the rules is compared with the sum of all the firing strengths. Therefore, the normalized firing strengths are computed in this layer as:

$$O_i^3 = \overline{w_i} = \frac{O_i^2}{\sum_i O_i^2} \quad (16)$$

Layer 4 implements the Sugeno-type inference system, i.e., a linear combination of the input variables of ANFIS, $x_1, x_2, \ldots x_p$ plus a constant term, $r_1, r_2, \ldots r_p$, form the output.

$$O_i^4 = y_i = \overline{w_i} f = \overline{w_i}(p_i x + r_i) \quad (17)$$

, where parameters $p_1, p_2, \ldots, p_i$ and $r_1, r_2, \ldots, r_i$, in this layer are referred to as the consequent parameters. In layer 5 we take:

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i \overline{w_i} f}{\sum_i \overline{w_i}} \quad (18)$$

In the last layer the consequent parameters can be solved for using a least square algorithm as:

$$Y = X \cdot \theta \quad (19)$$

, where X is the matrix

$$X = [w_1 x + w_1 + w_2 x + w_2 + w_3 x + w_3] \quad (20)$$

,where x is the matrix of inputs for the nonlinear part of the STAR models with fuzzy rules. The next step is to take also into the process the autoregressive linear part and the matrix of (20) becomes:

$$X = [t + y_{t-1} + w_1 x + w_1 + \ldots. + w_{\cdot 3} x + w_3] \quad (21)$$

,where t is taken for the constant estimation of the autoregressive linear part of STAR models and it is a vector of ones, while y_{t-1} is the dependent variable with one lag for the linear part. In (19) θ is a vector of unknown parameters as:

$$\theta = \left[\pi_{10}, \pi_{11}, p_1, q_1, r_1, p_2, q_2, r_2, ..., p_9, q_9, r_9\right]^T \quad (22)$$

,where p,q and r are the consequent parameters for the nonlinear pat of STAR models, which are the usual parameters in neuro-fuzzy procedure, while also we have some additional parameters, π_{10}, π_{11}, which are the estimated coefficients of the linear part and T indicates the transpose. Because the normal least square method leads to singular inverted matrix we use the singular value decomposition (SVD) with Moore-Penrose pseudoinverse of matrix [13]-[15]. So we observe that we have added some extra parameters in the neuro-fuzzy approach following the STAR modelling. In the case of one input and three linguistic terms each rule has one parameters and plus the constant there will be 6 parameters for the nonlinear part of STAR models, while there will be two parameters, the constant and the AR(1) for the autoregressive linear part. Similarly, for two inputs with three linguistic terms, which means that we have an AR(2) process, the rules are

IF y_{t-1} *is low AND* y_{t-2} *is low THEN* $f_1=p_1x_1 + q_1x_2 + r_1$

IF y_{t-1} *is low AND* y_{t-2} *is medium THEN* $f_2=p_2x_1 + q_2x_2 + r_2$

IF y_{t-1} *is low OR* y_{t-2} *is high THEN* $f_3=p_3x_1 + q_3x_2 + r_3$

IF y_{t-1} *is medium AND* y_{t-2} *is low THEN* $f_4=p_4x_1 + q_4x_2 + r_4$

IF y_{t-1} *is medium AND* y_{t-2} *is medium THEN* $f_5=p_5x_1 + q_5x_2 + r_5$

IF y_{t-1} *is medium AND* y_{t-2} *is high THEN* $f_6=p_6x_1 + q_6x_2 + r_6$

IF y_{t-1} *is high AND* y_{t-2} *is low THEN* $f_7=p_7x_1 + q_7x_2 + r_7$

IF y_{t-1} *is high AND* y_{t-2} *is medium THEN* $f_8=p_8x_1 + q_8x_2 + r_8$

IF y_{t-1} *is high AND* y_{t-2} *is high THEN* $f_9=p_9x_1 + q_9x_2 + r_9$

Then the steps for ANFIS computation will be:

In the first layer we generate the membership grades

$$O_i^1 = \mu_{A_i}(x_1),\ \mu_{B_i}(x_2) \quad (23)$$

, where x_1 and x_2 are the inputs (the lagged dependent variable of lag order one and two). In layer 2 we generate the firing strengths or weights:

$$\begin{aligned} O_i^2 &= w_i = \prod_{j=1}^{m} (\mu_{A_i}(x_1), \mu_{B_i}(x_2)) = \\ &andmethod\,(\mu_{A_i}(x_1), \mu_{B_i}(x_2)) \\ &= product\,(\mu_{A_i}(x_1) * \mu_{B_i}(x_2)) \end{aligned} \quad (24)$$

In layer 2 we use the *AND* relation, so we take the *product* operator. In layer 3 we normalize the firing strengths. Because we have nine rules will be:

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2 + w_3 + ... + w_8 + w_9} \quad (25)$$

, where i is for $i=1,2...,9$. In layer 4 we **c**alculate rule outputs based on the consequent parameters.

$$O_i^4 = y_i = \overline{w_i} f_i = \overline{w_i}(p_i x_1 + q_i x_2 + r_i) \quad (26)$$

Layer 5 remains the same, while X input matrix will be:

$$X = [t + y_{t-1} + y_{t-2} + w_1 x + w_1 + ... + w_9 x + w_9] \quad (27)$$

, where the inputs are defined as previously, while y_{t-2} is the dependent variable with two lags and vector θ will be:

$$\theta = \left[\pi_{10}, \pi_{11}, \pi_{12}, p_1, q_1, r_1, p_2, q_2, r_2, ..., p_9, q_9, r_9\right]^T \quad (28)$$

For the first layer and (14), (23) we use the logistic and exponential membership functions as we have defined them in (6)-(7). In order to find the optimized antecedent parameters we use backpropagation algorithm [16]-[18]. The parameter update for example of the parameter c in (6)-(7) is:

$$c_{ij}(n+1) = c_{ij}(n) - \frac{\eta_c}{p} \cdot \frac{\partial E}{\partial c_{ij}} \quad (29)$$

, where η_x is the learning rate for the parameter c_{ij}, p is the number of patterns and E is the error functions which is

$$E = \frac{1}{2}(y - y^t)^2 \quad (30)$$

, where y^t is the target-actual and y is ANFIS output variable. The chain rule in order to calculate the derivatives used to update the membership function parameters are
:

$$\frac{\partial E}{\partial c_{ij}} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial y_i} \cdot \frac{\partial y_i}{\partial w_i} \cdot \frac{\partial w_i}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial c_{ij}} \quad (31)$$

The partial derivatives for two inputs are derived below:

$$\frac{\partial E}{\partial y_{ij}} = y - y^{t} = e \tag{32}$$

For the output is

$$y = \sum_{i=1}^{n} y_i \tag{33}$$

, hence it will be

$$\frac{\partial y}{\partial y_i} = 1 \tag{34}$$

$$y_i = \frac{w_i}{\sum_{i=1}^{n} w_i}(p_i x_1 + q_i x_2 + r_i) \tag{35}$$

, hence it will be

$$\frac{\partial y_i}{\partial w_i} = \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^{n} w_i} \tag{36}$$

$$w_i = \prod_{j=1}^{m} \mu_{A_{ji}} \tag{37}$$

, and it will be

$$\frac{\partial w_i}{\partial \mu_{ji}} = \frac{w_i}{\mu_{ji}} \tag{38}$$

The last partial derivative, $\frac{\partial \mu_{ij}}{\partial c_{ij}}$, depends on the membership function we examine. The update equations for antecedent $c_{ij,}$ and γ_{ij} parameters of exponential function are:

$$c_{ij}(n+1) = c_{ij}(n) - \eta_{\mathrm{c}} \cdot e \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \left(\frac{2\gamma(\mathrm{x}_{ij} - \mathrm{c}_{ij})\exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})^2}{(1 - \exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})^2\ ))^2} \right) \tag{39}$$

$$\gamma_{ij}(n+1) = \gamma_{ij}(n) - \eta_{\gamma} \cdot e \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \left(\frac{(\mathrm{x}_{ij} - \mathrm{c}_{ij})^2 \exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})^2}{(1 - \exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})^2))^2} \right) \tag{40}$$

The update equations for antecedent $c_{ij,}$ and γ_{ij} parameters of logistic function are:

$$c_{ij}(n+1) = c_{ij}(n) - \eta_{\mathrm{c}} \cdot e \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \left(\frac{-\gamma(\mathrm{x}_{ij} - \mathrm{c}_{ij})\exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij}))}{(1 + \exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})\ )\ )^2} \right) \tag{41}$$

$$\gamma_{ij}(n+1) = \gamma_{ij}(n) - \eta_{\gamma} \cdot e \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \left(\frac{(\mathrm{x}_{ij} - \mathrm{c}_{ij})\exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij}))}{(1 + \exp(-\gamma_{ij}(\mathrm{x}_{ij} - \mathrm{c}_{ij})\ )\ )^2} \right) \tag{42}$$

The next step is to take the initial values for center and bases parameters, or c and γ parameters. To be specific we define the samples, for example for one input, such as in Table I.

TABLE I
SAMPLES FOR INITIAL VALUES

Low	Medium	High
If y_{t-1}>min(y_{t-1}) and if y_{t-1}<mean(y_{t-1})	If y_{t-1}>mean(y_{t-1}) and if y_{t-1}<Q3(y_{t-1})	If y_{t-1}>Q3(y_{t-1}) and if y_{t-1}<max(y_{t-1})

, where y_{t-1} is defined as previously, min, mean, Q3 and max denote respectively, the minimum, the average, the third quartile and the maximum values of inputs. Based on these values we take the average of each sample in each linguistic term as the initial value for parameter c. For parameter γ we have two options. The first one is to take the standard deviations of the samples of Table I, while the second option is to take as initial values 1.5 in all cases. We follow the second option. The learning rates for all antecedent parameters are set up at 0.5 and the number of maximum epochs at 100.

The forecasting performance of STAR models in both in-sample and out-of- sample periods is counted based on the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) described respectively by (43) and (44).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \qquad (43)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (44)$$

III. DATA

We examine two macroeconomic variables of US economy, the inflation rate and the six monthly treasury bills interest rates. The data we use in our analysis are in monthly frequency. We examine the period 1950-2009. The period 1950 to 2005 is used for the in-sample period and the period 2006-2009, which is a period 48 observations, is left for the out-of-sample forecasting period.

IV. EMPIRICAL RESULTS

In Table II we present the results of ADF and KPSS tests, The results are mixed. For gross domestic product we reject unit root in α=0.05 and 0.10 based on ADF test, while we accept stationarity only in α=0.01 based on KPSS test. For inflation rates we reject unit root based on ADF statistic in all statistical significance levels, but we reject stationarity hypothesis based on KPSS test. We accept that treasury bills interest rates are stationary in first differences*, I(1),* based on both ADF and KPSS tests.

In Table III we report the linearity tests for the two macroeconomic variables we examine. The value of lag order *p* is chosen based on the minimum *p-value* and in the cases where there are more than one zero *p-values* lag order *p* is chosen based on the highest *F-statistic.* In all cases we found an autoregressive process AR(1), *p=1*. Based on this process we choose the lag order of delay 1 and 2 respectively for inflation and interest rates.

TABLE II
ADF AND KPSS TESTS

Indices	ADF-statistic	KPSS-statistic
Inflation Rate Levels	-5.040	0.3759
Inflation Rate First differences		0.1102
Treasury Bills Levels	-2.089	0.5547
Treasury Bills First differences	-8.461	0.0335
Critical values for ADF[1]	-4.086	α=0.01
	-3.471	α=0.01
	-3.162	α=0.10
Critical values for KPSS[2]	0.216	α=0.01
	0.146	α=0.01
	0.119	α=0.10

1 MacKinnon [19], 2 Kwiatkowski *et al.*,[9]

TABLE III
LINEARITY TESTS FOR INFLATION AND INTEREST RATES

Indices	Inflation Rate	Treasury Bills
p	1	1
d=1	24.820 (0.000)	7.503 (0.0001)
d=2	13.458 (0.000)	17.551 (0.000)
d=3	3.470 (0.0160)	5.941 (0.0001)
d=4	6.278 (0.0003)	3.069 (0.0274)
d=5	16.085 (0.000)	5.228 (0.0014)

*p-values in parenthese

In Table IV the estimated results for inflation and interest rates respectively with nonlinear squares and Levenberg-Marquardt algorithm, are reported. We observe, in the case of the inflation rate, that the fuzzy membership function parameters, *c* and *γ* are statistically significant. On the other hand for Treasury bill interest rates, parameter *γ* is statistically insignificant in the case of logistic function, while parameter *c* is significant in both cases.

In Table V we presents the estimated results of Neuro-Fuzzy STAR models for inflation and interest rates with one input, the dependent variable with one lag, while in Table VI we present the Neuro-Fuzzy STAR results with exponential membership function for treasury bill interest rates and with two inputs as for example. Similarly the estimations with logistic functions as well as for inflation rate can be derived.

For example we examine the six-monthly treasure bills interest rates with exponential neuro-fuzzy and we have the rules:

IF y_{t-1} is high or expansive THEN $f_1 = 0.1180\, y_{t-1} + 0.4961$
($\gamma_1 = 1.500$ and $c_1=5.500$)

IF y_{t-1} is medium THEN $f_2 = 0.2611\, y_{t-1} - 0.2604$
($\gamma_2 = 3.4998$ and $c_2=4.4671$)

IF y_{t-1} is low or recessive THEN $f_3 = 0.3604\, y_{t-1} - 0.2570$
($\gamma_3 = 1.500$ and $c_3=5.500$)

In any case is

$y_t = -0.0213 + 0.7395\, y_{t-1}$

, which is nothing else than the linear autoregressive part of the ANFIS-STAR model. A similar derivation of rules and estimations can be made for the other estimations of Tables V and VI

TABLE IV
NONLINEAR SQUARES ESTIMATIONS FOR INFLATION RATE TREASURY AND BILLS INTEREST RATES

	Linear Part		Non-Linear Part	
Exponential STAR	π_{10}	π_{11}	π_{20}	π_{21}
Inflation rate	0.0509 (1.018)	0.5971 (3.639)*	-0.1005 (-0.378)	-2.0167 (-4.506)*
	γ	c		
	2.9327 (1.997)**	0.1163 (1.671)***		
Logistic STAR	π_{10}	π_{11}	π_{20}	π_{21}
Inflation rate	0.2611 (2.173)**	-0.2701 (-0.639)	-0.3606 (-2.220)**	0.0317 (0.067)
	γ	c		
	4.875 (2.034)**	-0.0696 (-2.723)*		
Exponential STAR	π_{10}	π_{11}	π_{20}	π_{21}
Treasury Bills	0.0424 (0.602)	0.7752 (1.691)***	-0.0506 (-0.658)	-1.1159 (-2.111)**
	γ	c		
	3.335 (1.761)***	0.601 (15.916)*		
Logistic STAR	π_{10}	π_{11}	π_{20}	π_{21}
Treasury Bills	0.0140 (0.683)	0.0022 (0.017)	-0.0489 (-1.226)	-0.5597 (-2.548)**
	γ	c		
	4.206 (0.022)	0.638 (3.682)*		

*,**,*** denotes significance in 0.01, 0.05 and 0.10 respectively, t-statistics in parentheses

In Table VI parameters π_{10}, π_{11}, π_{12} are the estimated autoregressive coefficients of linear part, while parameters p_i, q_i and r_i for $i=1,2..9$, are the consequent parameters or the estimated coefficients of the nonlinear part of exponential STAR model. Parameters c_1, c_2 and c_3 are the center parameters of exponential function for linguistic terms low, medium and high respectively and for the first input, while parameters c_4, c_5 and c_6 are the respective centers for the second input. Similarly we have parameters γ.

In Table VII the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are reported. The values for ANFIS are refereed in one input and the estimations of Table V, while the forecasts with two inputs are very close. The results support Neuro-Fuzzy STAR. More specifically Neuro-Fuzzy STAR outperform significant the conventional STAR econometric modelling based on RMSE and MAE values in both in-sample and out-of-sample periods. Especially in the out-of-sample or testing period, which is of greater interest, the RMSE and MAE values are significant lower. This indicates that the ANFIS technology is more appropriate for STAR modelling. This can be explained by the fact that conventional STAR models do not consider linguistic terms, fuzzy rules and operators, as also the combination of fuzzy logic and neural networks with error backpropagation algorithm can be superior to the nonlinear squares estimation procedure.

TABLE V
NEURO-FUZZY ESTIMATIONS FOR INFLATION RATE TREASURY AND BILLS INTEREST RATES WITH ONE INPUT

	Linear Part		Non-Linear Fuzzy Part	
Exponential Neuro-Fuzzy STAR	π_{10}	π_{11}	p_1	p_2
Inflation rate	0.1259	0.5125	1.5429	0.1308
	p_3	r_1	r_2	r_3
	-0.1308	1.0305	1.834	-1.823
	c_1	c_2	c_3	
	6.800	3.3080	1.2438	
	γ_1	γ_2	γ_3	
	1.6676	2.5533	2.2386	
Logistic Neuro-Fuzzy STAR	π_{10}	π_{11}	p_1	p_2
Inflation rate	0.5990	-0.8477	1.4136	1.7482
	p_3	r_1	r_2	r_3
	-1.8944	0.5099	0.4012	-0.4601
	c_1	c_2	c_3	
	0.3566	0.0284	0.0002	
	γ_1	γ_2	γ_3	
	1.500	1.5128	1.5072	
Exponential Neuro-Fuzzy STAR	π_{10}	π_{11}	p_1	p_2
Treasury Bills	-0.0213	0.7395	0.1180	0.2611
	p_3	r_1	r_2	r_3
	0.3604	0.4961	-0.2604	-0.2570
	c_1	c_2	c_3	
	5.500	4.4671	1.6375	
	γ_1	γ_2	γ_3	
	1.500	3.4998	1.5853	
Logistic Neuro-Fuzzy STAR	π_{10}	π_{11}	p_1	p_2
Treasury Bills	0.5510	0.7421	-0.1139	0.6228
	p_3	r_1	r_2	r_3
	0.2331	1.7483	-0.4788	-0.7185
	c_1	c_2	c_3	
	5.500	1.3802	-9.9824	
	γ_1	γ_2	γ_3	
	1.500	1.2318	-0.1619	

TABLE VI
NEURO-FUZZY EXPONENTIAL-STAR ESTIMATIONS FOR TREASURY AND BILLS INTEREST RATES WITH TWO INPUTS

Linear Part		Non-Linear Fuzzy Part	
π_{10}	π_{11}	π_{12}	p_1
0.8656	-0.7134	-0.5307	-1.4814
p_2	p_3	p_4	p_5
0.9164	1.1928	5.2625	-0.2805
p_6	p_7	p_8	p_9
-0.4900	0.4561	0.5388	-0.3797
q_1	q_2	q_3	q_4
0.0968	-1.5012	1.8325	2.4020

q_5	q_6	q_7	q_8
-3.0680	-5.8487	1.4488	-0.9066
q_9	r_1	r_2	r_3
-1.2020	0.1394	0.8174	1.1808
r_4	r_5	r_6	r_7
0.1103	0.1242	-0.1799	-0.1213
r_8	r_9	c_1	c_2
0.5025	-0.6214	-1.2691	0.2527
c_3	c_4	c_5	c_6
-0.6663	0.9921	2.2248	-0.5928
γ_1	γ_2	γ_3	γ_4
1.7305	1.8732	0.8733	1.7466
γ_5	γ_6		
2.1519	0.7738		

TABLE VII
RMSE AMD MAE VALUES

In sample period				
	Inflation rate		Interest rates	
	RMSE	MAE	RMSE	MAE
Exponential	0.2719	0.2057	0.1641	0.1153
Logistic	0.2696	0.1981	0.1643	0.1119
Neuro-Fuzzy Exponential	0.2255	0.1763	0.1440	0.1021
Neuro-Fuzzy Logistic	0.2040	0.1430	0.1480	0.1074
Out-of-sample period				
	Inflation rate		Interest rates	
	RMSE	MAE	RMSE	MAE
Exponential	0.2791	0.2240	0.3249	0.2741
Logistic	0.2857	0.2363	0.2929	0.2333
Neuro-Fuzzy Exponential	0.2619	0.2105	0.2263	0.1553
Neuro-Fuzzy Logistic	0.2699	0.2152	0.2161	0.1432

V. CONCLUSIONS

In this paper we proposed a neuro-fuzzy approach with error backpropagation optimization for STAR modeling. The reason we followed this procedure is because the nonlinear part of STAR models accounts for the membership grades of inputs but no rules or linguistic terms are included in the conventional econometric modeling. More over additional membership function can be proposed, as the triangular, trapezoidal, or Generalized Bell functions among others. Finally, other optimization methods can be applied in order to find the fuzzy parameters as the genetic algorithms instead to error backpropagation we have used in this study.

REFERENCES

[1] J. L. M. Aznarte, J. M. Benitez and J. L. Castro, “Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences”, *Fuzzy Sets and Systems,* vol. 158, pp. 2734-2745, 2007
[2] Atsalakis, S.G. and G.I. Atsalakis, “Fruit production forecasting by neuro-fuzzy techniques,” Paper prepared for presentation at the 113th EAAE Seminar “*A resilient European food industry and food chain in a challenging world*”, Chania, Crete, Greece, September, pp. 3-6, 2009
[3] Ballini, R., I. Luna, L.M. Lima and R.L.F. da Silveira, “A comparative analysis of neurofuzzy, ANN and ARIMA models for Brazilian stock index forecasting,” *SCE - Computing in Economics and Finance,* 1995
[4] D. Marček, “Stock Price Forecasting: Autoregressive Modelling and Fuzzy Neural Network,” *Proceedings of the 1999 EUSFLAT-ESTYLF Joint Conference*, Palma de Mallorca, Spain, September, pp. 22-25, 1995
[5] K.S. Chan and H. Tong, “On estimating thresholds in autoregressive models”, *Journal of Time Series Analysis*, vol. 7, pp. 178-190, 1986
[6] T. Teräsvirta, “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models”, *Journal of the American Statistical Association*, vol. 89, no. 425, 208–218, 1994
[7] T. Teräsvirta and H.M. Anderson, “Characterizing nonlinearities in business cycles using smooth transition autoregressive models”, *Journal of Applied Econometrics*, vol. 7, pp. 119-136, 1992
[8] D. A. Dickey, and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”, *Journal of the American Statistical Association*. vol. 74, pp. 427–431, 1979
[9] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root”, *Journal of Econometrics,* vol. 54, pp. 159-178, 1992
[10] M. S. Bartlett, “Periodogram analysis and continuous spectra,” *Biometrika,* vol. 37, pp. 1-16, 1950
[11] J.-S.R. Jang, “ANFIS: Adaptive-Network-based Fuzzy Inference Systems,” *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 23, No. 3, pp. 665-685, 1993
[12] J.-S. R. Jang and C.-T. Sun, “Neuro-fuzzy Modeling and Control,” *Proceedings of the IEEE*, vol. 83, no. 3, 378-406, March, 1995
[13] E.H. Moore, “On the reciprocal of the general algebraic matrix,” *Bulletin of the American Mathematical Society,* vol. 26, pp. 394–395, 1920
[14] R. Penrose, “A generalized inverse for matrices,” *Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 406–413, 1955
[15] M. Petrou, and P. Bosdogianni, *Image Processing: The Fundamentals*, John Wile & Sons, 2000
[16] L. Khan, S. Anjum and R. Bada, “Standard Fuzzy Model Identification using Gradient Methods,” *World Applied Sciences Journal*, vol. 8, No. 1, pp. 01-09, 2010
[17] M.A. Denai, F. Palis and A. Zeghbib, ANFIS Based Modelling and Control of Non-Linear systems: A Tutorial, *IEEE Conf. on Systems, Man, and Cybernetics*, vol. 4, pp. 3433-3438, 2004
[18] L.H. Tsoukalas, and R.E. Uhrig, *Fuzzy and Neural Approaches in Engineering*, First Edition, John Wiley & Sons, 1997, pp. 445-470
[19] J. G. MacKinnon, “Numerical Distribution Functions for Unit Root and Cointegration Tests”, *Journal of Applied Econometrics*, vol. 11, pp. 601-618, 1996